

MODEL-BASED STRATIFICATIONS FOR ENHANCING THE DETECTION OF RARE ECOLOGICAL EVENTS

THOMAS C. EDWARDS, JR.,^{1,6} D. RICHARD CUTLER,² NIKLAUS E. ZIMMERMANN,³ LINDA GEISER,⁴
AND JIM ALEGRIA⁵

¹U. S. Geological Survey Biological Resources Division, Utah Cooperative Fish and Wildlife Research Unit,
Utah State University, Logan, Utah 84322-5290 USA

²Department of Mathematics and Statistics, Utah State University, Logan, Utah 84322-3900 USA

³Department of Landscape Research, Swiss Federal Research Institute WSL, Zuercherstrasse 111,
CH-8903 Birmensdorf, Switzerland

⁴USDA Forest Service, Siuslaw National Forest, P.O. Box 1148, Corvallis, Oregon 97339 USA

⁵Bureau of Land Management Oregon State Office, 333 SW First Avenue, Portland, Oregon 97204 USA

Abstract. A common concern when designing surveys for rare species is ensuring sufficient detections for analytical purposes, such as estimating frequency on the landscape or modeling habitat relationships. Strict design-based approaches provide the least biased estimates but often result in low detection rates of rare species. Here, we demonstrate how model-based stratification can improve the probability of detecting five rare epiphytic macrolichens (*Nephroma laevigatum*, *N. occultum*, *N. parile*, *Lobaria scrobiculata*, and *Psuedocyphelaria rainierensis*) in the Pacific Northwest. We constructed classification tree models for four more common lichens (*L. oregana*, *L. pulmonaria*, *P. anomala*, and *P. anthraspis*) that are associated with the rare species, then used the models to generate strata for sampling for the five lichen species considered rare. The classification tree models were developed using topographic and bio-climatic variables hypothesized to have direct relationships to the presence of the modeled lichen species. When the expected detection rates using the model-based stratification approach was tested on an independent data set, it resulted in two- to fivefold gains in detection compared to the observed detection rates for four of the five tested rare species.

Key words: classification trees; epiphytic lichens; model-based stratification; Northwest Forest Plan; rare lichens; sampling; species rarity.

INTRODUCTION

Ecologists and conservationists often deal with rare or uncommon species. Reasons for interest in these species range from curiosity about the underlying ecological relationships that determine rarity (Goerck 1997) to the need for information about the species so that proper management and conservation strategies can be developed (Sheldon 1988, Maina and Howe 2000). A common concern when designing surveys for rare species is ensuring sufficient detections for analytical purposes (Green and Young 1993, Edwards et al. 2004), be they for estimations of frequency on landscapes (Alexander et al. 1997), or for use in explanatory or predictive models of suitable habitat or spatial distribution (Hill and Keddy 1992, Wiser et al. 1998).

One specific example where the species of interest appear to be ecologically rare, and for which there exists a strong need for estimates of frequencies for conservation purposes, is in the U.S. Pacific Northwest and a subset of that area defined by the Pacific North-

west Forest Management Plan (hereafter NWFP). The NWFP was a direct consequence of a court-mandated requirement to develop a scientifically credible conservation strategy for the Northern Spotted Owl *Strix occidentalis* (see Thomas et al. 1990, Murphy and Noon 1992) while also maintaining biodiversity on public lands within the area of the Plan. The Record of Decision leading to the Plan (ROD; U.S. Department of Agriculture, Forest Service, and U.S. Department of Interior, Bureau of Land Management 1994), and subsequent amendments (U.S. Department of Agriculture, Forest Service, and U.S. Department of Interior, Bureau of Land Management 2001), identify more than 350 species of concern in the Plan area that required management plans. These species included fungi, lichens, bryophytes, terrestrial mollusks, arthropods, a small number of vascular plants, and several terrestrial vertebrates. Collectively, these species and species groups were referred to as Survey and Manage species. All were hypothesized to have associations with old forests, and concern existed over the possible impacts of forest management on the persistence of these species.

One immediate realization was that insufficient information existed to ascertain the impacts of conservation decisions on Survey and Manage species. Most

Manuscript received 1 April 2004; revised 29 June 2004; accepted 1 July 2004; final version received 13 September 2004.
Corresponding Editor: A. M. Ellison. For reprints of this Special Feature, see footnote 1, p. 1079.

⁶ E-mail: tce@nr.usu.edu

of the species were known from relatively few sites, and their spatial distributions were unclear as well. Even less was known about basic life history attributes. In one sense, the Survey and Manage species are rare events in time and space. They are often associated with microsite characteristics that are themselves rare and patchy in time and space (lichens [McCune 1993, Esseen et al. 1996]). Many also exhibit irruptive population behavior, appearing only for brief periods or only once or twice over extended periods (fungi [Molina et al. 2001]), while in other cases, large distances may exist between individuals or populations (fungi [Dreisbach et al. 2002]). These characteristics make it difficult to design statistically valid sampling schemes, particularly when little is known about the species' distribution or habitat requirements that could be used to focus sampling efforts.

Sound study designs and analytical techniques for estimating species frequencies are well represented in the scientific literature (see Schreuder et al. 1993, Olsen and Schreuder 1997) and have been successfully applied to diverse resource issues (forest resources [Gregoire et al. 1995, Shiver and Borders 1996], animals [Sauer and Droegge 1990, James et al. 1996], aquatic systems [Stow et al. 1998]). However, many of these approaches rely on design-based sampling efforts. While design-based approaches clearly provide the least biased estimates, they unfortunately are often of insufficient sampling intensity to ensure enough observations of rare species. Methods such as adaptive cluster sampling (Lohr 1999, Cutler et al. 2002, Philippi 2005) can increase sampling efficiencies, but themselves can be limited by the information needed for cluster starting points, or the need for immediate identification of the sample unit of interest so that the cluster can be identified and additional samples collected or measured.

One way to increase detection probabilities for rare species is to use presence-absence information from other, more abundant but related species to generate strata for sampling for the rare species. Strata could be derived from any number of available statistical models capable of relating a binary response to a set of predictor variables, such as generalized additive models, logistic regression, and classification trees (see Hoeting et al. 2000). Here, we examine the efficacy of a model-based stratification approach based on classification trees to enhance detection probabilities for rare epiphytic macrolichens in the Pacific Northwest. We fit classification trees for four common lichen species, using selected topographic and bio-climatic variables all hypothesized to have direct relationships to the presence of the common lichen species. The fitted trees were then used to define sampling strata for associated, rare lichen species. Specifically, we test whether model-based strata developed from more common, associated species can improve the detection probabilities of rare species. We next evaluate the efficacy of the

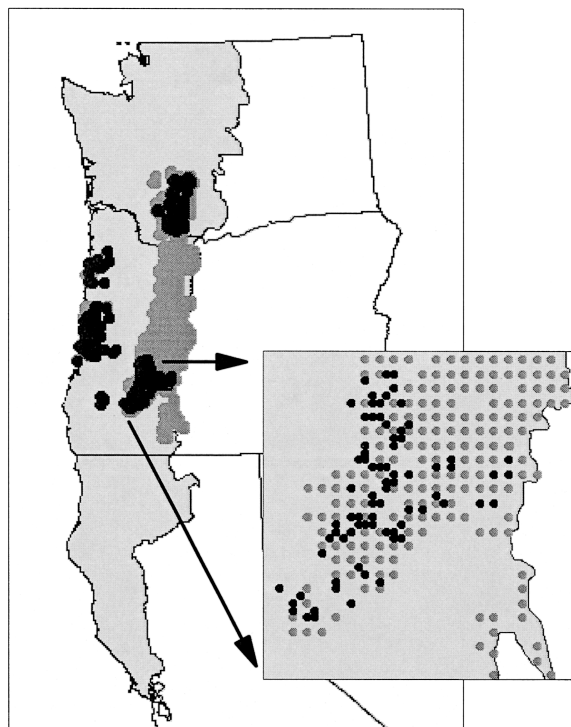


FIG. 1. Generalized location of the LAQ (lichen air quality) model building plots (medium gray) and PILOT validation plots (black) in relation to the area of the Northwest Forest Plan (light gray). The figure inset is a representation of the validation PILOT points (black) and LAQ model building points (medium gray).

stratification process for predicting occurrences of rare species on an independent data set within the same geographic area, comparing the gains in the model-based detection probabilities against those realized from a systematic random sample design.

METHODS

Study area and species

Data used in our analyses were collected in the NWFP area of the U.S. Pacific Northwest (Fig. 1). Two subset study areas within the plan boundary were used in our analyses. The first (hereafter LAQ) was used for model building purposes and included sampling sites on seven national forests: the Umpqua, Willamette, Mt. Hood, Deshutes, Gifford-Pinchot, Siuslaw, and Winema (Fig. 1). Epiphytic macrolichen species were surveyed from 1993 to 2000 as part of a larger effort using lichens to evaluate air quality in the Pacific Northwest (Geiser 2004). The second survey (hereafter PILOT) was conducted to provide data on rare, old-forest-associated species to the NWFP Survey and Manage program. These data served as an independent assessment tool for the models. The PILOT surveys included forests in three regions: the southern Washington Cascades (Gifford Pinchot National Forest); the Oregon Coast Range (Siuslaw National Forest and portions of

TABLE 1. Numbers (percentage) of detections of rare species that are co-located with more common species in the LAQ (lichen air quality) study area within the range of the Pacific Northwest Forest Plan.

Rare species	Common species†			
	<i>Lobaria oregana</i>	<i>Lobaria pulmonaria</i>	<i>Pseudocyphellaria anomala</i>	<i>Pseudocyphellaria anthraspis</i>
<i>Lobaria scrobiculata</i>		37/47 (78.7)	39/47 (83.0)	
<i>Nephroma laevigatum</i>		24/25 (96.0)	19/25 (76.0)	19/25 (76.0)
<i>Nephroma occultum</i>		10/13 (76.9)	13/13 (100.0)	
<i>Nephroma parile</i>		24/31 (77.4)	27/31 (87.1)	
<i>Pseudocyphellaria rainierensis</i>	8/9 (88.9)	8/9 (88.9)	7/9 (77.8)	

Note: Empty cells indicate species that were never co-located.

† Numerator = number of detections of the rare species that are co-located with the common species; denominator = total number of occurrences of the rare species in the LAQ study area.

Coos Bay, Eugene, and Salem Bureau of Land Management lands); and the Umpqua Basin (Umpqua National Forest, Roseburg BLM [Bureau of Land Management], and portions of the Willamette National Forest, and Eugene and Coos Bay BLM).

All sample sites in both the LAQ and PILOT study areas were surveyed at least once and samples of each detected epiphytic macrolichen were collected for subsequent identification. Crustose (e.g., *Calicium* spp. and *Chaenotheca* spp.) and ground-dwelling species (e.g., *Cladonia norvegica*, *Peltigera pacifica*) were not surveyed at all sites and were excluded from analysis. A total of 49 lichen species having designations ranging from rare to common were considered. The four most abundant species in the LAQ surveys that were also sampled in the PILOT surveys, *Lobaria oregana*, *L. pulmonaria*, *Pseudocyphellaria anomala*, and *P. anthraspis*, were used for modeling purposes. Each of these species was detected on >120 of the 840 LAQ sites, providing sufficient sample sizes for developing model-based strata. Cross-tabulations of these species with other, rarer species detected in both the LAQ and PILOT surveys revealed substantial overlap of five specific rare species with the four identified more common species (Table 1). Accordingly, we fit classification trees for each of these four common species, with a view to using the resulting predictive model for *L. pulmonaria* to predict the presences of *Nephroma laevigatum*, *L. scrobiculata*, and *P. rainierensis*; the models for *L. oregana* and *P. anthraspis* to predict presences of *P. rainierensis*; and the model for *P. anomala* to predict presences for *N. occultum*, *L. scrobiculata*, and *N. parile*.

Data structure and characteristics

Data in both study areas were collected on the current vegetation survey plots (CVS), a systematic grid overlaid on all Forest Service and BLM lands in the Pacific Northwest. Its principal application is the generation of estimates of forest resources (see Max et al. 1996). A total of 840 CVS plots were sampled in the LAQ, and an additional 300 sites in the PILOT study area. Sample sites were apportioned equally among the three areas comprising the PILOT study area. Presence and

absence of each lichen species was recorded on a 1-acre (0.4-ha) plot centered on the central (#1) subplot on each CVS site for the LAQ survey (details in Edwards et al. 2004). Plot size for the PILOT surveys was 0.5 acres (0.2 ha). If the purpose of our study was to compare the estimated percent occupancy rates from the LAQ and PILOT surveys, the difference in the size of the sample units would be a concern given that larger plots will have higher probabilities of occupancy. However, the purpose of our analyses is to use the LAQ data to fit a model that can be used to predict PILOT survey sites that are more or less likely to be occupied by the rare lichens, with a view to developing a stratification for the PILOT sampling. For this application, it does not matter if the PILOT plot size is the same, larger, or smaller than the LAQ plot size.

All plot locations were intersected with spatially explicit maps of topographic and bio-climatic variables (Table 2) in a geographic information system (GIS). The selected topographic and bio-climatic variables were all hypothesized to have direct relationships to the presence of the modeled lichen species. Ninety-meter resolution topographic variables (slope, aspect and elevation) were obtained by resampling the 30-m resolution national elevation data set (NED) (Gesch et al. 2002). Bio-climatic variables were derived from the DAYMET 1-km daily gridded weather surfaces that have been reduced to 18-yr monthly and yearly climatological summaries (1981–1998; data is from DAYMET U.S. Data Center for Daily Surface Weather Data and Climatological Summaries, available online).⁷

DAYMET generates daily surfaces of temperature, precipitation, humidity, and radiation over large areas of complex terrain (Thornton et al. 1997, Thornton and Running 1999). It uses digital topographic models and observations of precipitation, maximum and minimum temperature from ground-based meteorological stations to generate other bio-climatic variables. Derivations for the modeled variables follow the DAYMET methodology described by Thornton et al. (1997), applied to monthly averages. Only the variable potential evapotranspiration (ETP), which is not part of the DAYMET

⁷ <http://www.daymet.org>

TABLE 2. Topographic and bio-climatic variables used to model the probability of presence for four common lichen species in the LAQ study area of the Pacific Northwest Forest Plan.

Variable type/ name	Description	Units
Topographic		
SLPE	percent slope	percent, 0–90
ASPE	aspect	degrees, 1–360
ELEV	elevation	m
Bio-climatic		
PREC	precipitation	cm
TMAX	maximum temperature	°C
TMIN	minimum temperature	°C
TAVE	monthly average temperature	°C
ETPJ	potential evapotranspiration	mm
RELH	relative humidity	percent
VPAM	ambient vapor pressure	Pa
VPAS	saturated vapor pressure	Pa
SFMM	monthly potential global radiation	kJ
TDAY	monthly average daytime temperature	°C
MIND	monthly moisture index	cm

Note: Classification tree models based on these variables were used to generate sampling strata for rare species.

procedures, was derived differently (see Jensen and Haise 1963). We down-scaled the 1-km DAYMET variables to 90-m resolution by first generating moving window regressions between the monthly averages of PREC, TMIN, and TMAX, and the topographic models, and then applying the regression parameters to the center cell of each 1-km window position. The regression parameters (i.e., lapse rates and intercept) were next interpolated to a 90-m resolution using inverse distance weighed interpolations, thus generating monthly 90-m resolution maps for each bio-climatic variable.

Preliminary analyses showed that correlations among the monthly values for the 11 sets of bio-climatic predictor variables were extremely high. Such extreme collinearity has implications for modeling. For example, two variables that essentially contain the same information can be selected in models for different species, implying differences in the models that are not real. The same phenomenon can occur even at different stages of the same classification tree model. To address the issue of collinearity, a principal components analysis was carried out on each of the 11 sets of monthly bio-climatic predictors.

In each case, the first principal component was an average of the 12 monthly measurements, while the second principal component was a contrast of values for six so-called summer months (April–September) to the six so-called winter months (October–March). For each set of 12 monthly variables, these two principal components explained over 95% of the variability, and in most cases the first two principal components explained over 99% of the variability in the sets of variables. Accordingly, for each set of monthly bio-climatic predictors we defined two new variables: (1) the average of the 12 monthly variables; and (2) the difference between the sum of the summer monthly values

and the winter monthly values, divided by 12. Hereafter we use the variable suffix “A” to denote the average of the 12 monthly measurements, and the suffix “D” to denote the difference derived variable. Thus, TMINA is the average minimum temperature for the 12 months and PRECD is the difference between summer and winter precipitation.

Statistical modeling and assessment

We used classification trees (Breiman et al. 1984) to relate the LAQ lichen presences to the modeled bio-climatic and topographic predictor variables. Classification trees have several advantages over other discrimination techniques (e.g., GLMs, GAMs), principal among these being the ability to deal with nonlinearities in the predictor variables and assess interactions, and ease of interpretation. For the most part we followed the approach of De'ath and Fabricius (2000), but instead of allowing the tree size to be determined completely by cross-validation and the 1-SE rule, we set the complexity parameter at 0.02 instead of using the default value of 0.01. This relaxation in complexity results in smaller, easier to interpret classification trees that have cross-validated error rates at least as good as the larger trees obtained from the default value. The classification trees were fit using the rpart library of functions in the R statistical package (Ihaka and Gentleman 1996 and online resources).⁸

We next fit the LAQ-based classification trees to our spatially explicit predictors within the GIS and modeled the probability of presence of each the four common lichen species across the entirety of the NWFP region. This was accomplished by applying the classification tree to the GIS layers and generating new maps that portrayed, for each 90-m pixel, the proba-

⁸ <http://www.r-project.org/>

bility of each of the four common lichen species being present at that particular pixel. Proportions of presences from sample plots in the PILOT study area were then examined against predicted probability for that plot, allowing us to link the presence or absence of each of the four common species to an estimate of the probability of presence. Ideal concordance between the LAQ models and the common species in the PILOT area would be represented as a 45° line on a scatter plot of observed probabilities of presence against predicted probabilities of presence (see Ferrier and Watson 1997: Fig. 6.2). Data and a description of the process used to estimate expected values for each rare species under different sampling probabilities are found in Appendices A–D.

Three measures of model error were used to evaluate model performance. These measures were: (1) a re-substitution (model) error rate, calculated using the LAQ data and hence providing an indication of how well the classification tree fit the modeling data; (2) a 10-fold cross-validation (Manly 1997) estimate of error obtained by using 1000 subsamples of the original LAQ data; and (3) a prediction error rate for the PILOT surveys, based on a probability of presence threshold of $P > 0.5$.

Under the Survey and Manage program, actual detections as opposed to region-wide estimates were of paramount importance. To reflect this need and desire for additional observed occurrences of the rare species, we chose to allocate sample units proportionally to the estimated probability of occurrence of the associated common species in each stratum. Different researchers with different objectives could use a different objective function for assigning sample units to the strata and still use all the other methodology described in this section.

RESULTS

Although the classification tree models are purely predictive in nature, it is interesting to note that the first split in each classification tree is on an average temperature index (Fig. 2) and that precipitation only occurs in one of the four classification trees (*L. oregana*). No topographic variables entered any of the models as significant predictors. Not unexpectedly, the LAQ model error estimates were lower than the LAQ cross-validated error (Table 3), ranging from a low of 10.2% to a high of 18.3%. Cross-validation error ranged from approximately 13% to 22%. Prediction errors were 15% to 27%, indicating that the LAQ models had, on average, about a one in four chance of being wrong. The cross-validated error rates were much closer to the observed error rates on the PILOT data, and likely better reflect the true error of the models.

LAQ models for the four common species, when evaluated as plots of the observed occurrences versus the predicted probabilities of occurrence, indicate all models had substantial predictive power (Fig. 3). Mod-

els for *L. pulmonaria*, *L. oregana*, and *P. anomala* are best, with the estimated values and their 95% confidence intervals encompassing the 45° line. The model for *P. anthraspis* is not as good, with observed values tailing off as the predicted probability of occurrence increases, indicating on over-prediction of potential locations for these species.

To evaluate the efficacy of the LAQ classification trees in determining a stratification for sampling of the rare, associated species, we calculated the estimated number of detections expected for each of the rare species had the predicted probabilities of presence obtained from the LAQ classification trees been used for stratification purposes. For almost all combinations of rare species and models of the more abundant species, the expected number of detections would have been higher under our model-based stratification than under the systematic design that characterized the PILOT surveys (Table 4). Gains in sampling efficiencies for four of the five species (*L. scrobiculata*, *N. laevigatum*, *N. occultum*, and *N. parile*) ranged from 1.2 to 5.0-fold, while results showed no difference for *P. rainierensis*. These increases in likely detections indicate that use of the LAQ models as the basis for developing sampling strata substantially increase the likelihood of detection of the rare lichens.

DISCUSSION

The models of the four common lichen species used to develop the prediction strata for the rare lichens all showed similar error rates when evaluated on the LAQ data set. However, when tested on the independent PILOT data set, the *P. anthraspis* and *P. anomala* models performed best, followed by *L. pulmonaria*. The *L. oregana* model produced the highest error rate. Overall, the models were successful in generating clear improvements in estimated, or expected, detection rates for four of the five rare lichen species (Table 4). Only the model of *L. oregana* applied to *P. rainierensis* did not perform well. The most likely explanation for the low predictive extrapolation of this last model is the lack of spatially explicit depictions of stand age for use in building the associative models. In this case, *L. oregana* and *P. rainierensis* are both associated with old forest, but *rainierensis* is more likely to be found in 300+-yr-old stands. Such subtle differences in ecological relationships like this could easily confound model-based stratifications, increasing the likelihood of prediction error. This observation reinforces the importance of having a solid understanding of species ecology when model building, as well as the importance of testing any predictive model with an independent test data.

Existing knowledge on lichen ecology suggests that prediction error rates may be reduced substantially by including additional information about forest stand characteristics (e.g., stand age, size classes, hardwood shrubs), and local moisture content (McCune 1993, Sil-

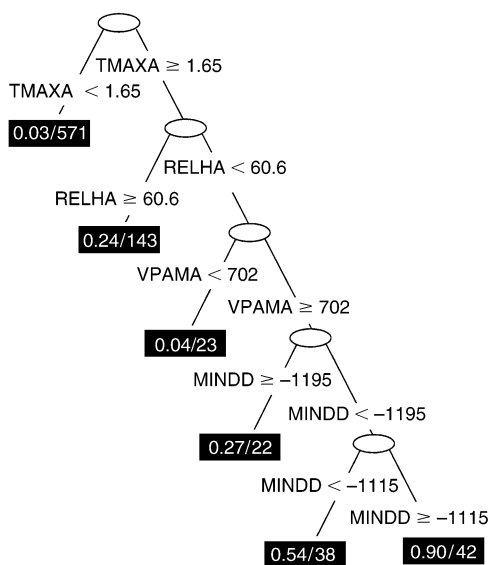
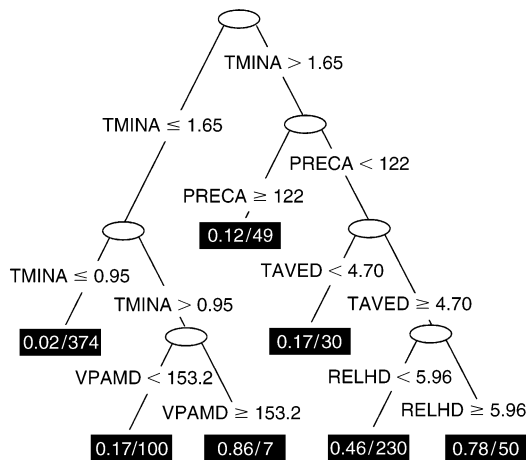
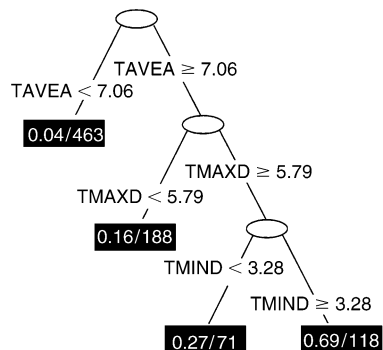
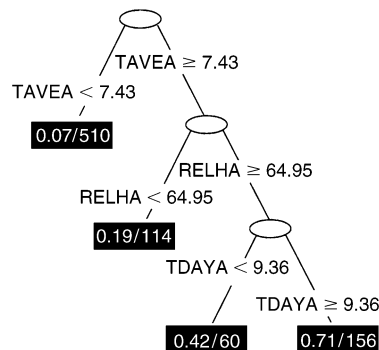
A) *Pseudocyphellaria anthraxis*B) *Lobaria oregana*C) *Pseudocyphellaria anomala*D) *Lobaria pulmonaria*

FIG. 2. Classification trees of the four modeled common lichen species used to develop prediction strata for the five rare lichen species in the area of the Northwest Forest Plan. See Table 2 for variable descriptions.

lett and Goslin 1999, Rosso et al. 2000, Peterson and McCune 2001) into models to be used for stratification. However, these types of variables were not available in spatially explicit formats for the entire study region, and hence could not be used for the purposes of strat-

ification in advance of sampling. Spatially explicit depictions of these kinds of variables are simply unavailable in the vast majority of predictive modeling efforts, often forcing models to be built using surrogates (e.g., topography) that may or may not have rel-

TABLE 3. Measures of error (%) for classification tree models built for four species of lichen in the LAQ study area, and assessed using independent data collected in the PILOT study area, Pacific Northwest Forest Plan.

Species	LAQ stratification models		PILOT assessment
	Model error	Cross-validation error	Prediction error
<i>Lobaria oregana</i>	18.3	22.5	27.3
<i>Lobaria pulmonaria</i>	15.2	18.3	19.3
<i>Pseudocyphellaria anomala</i>	12.6	15.4	15.0
<i>Pseudocyphellaria anthraxis</i>	10.2	13.2	15.3

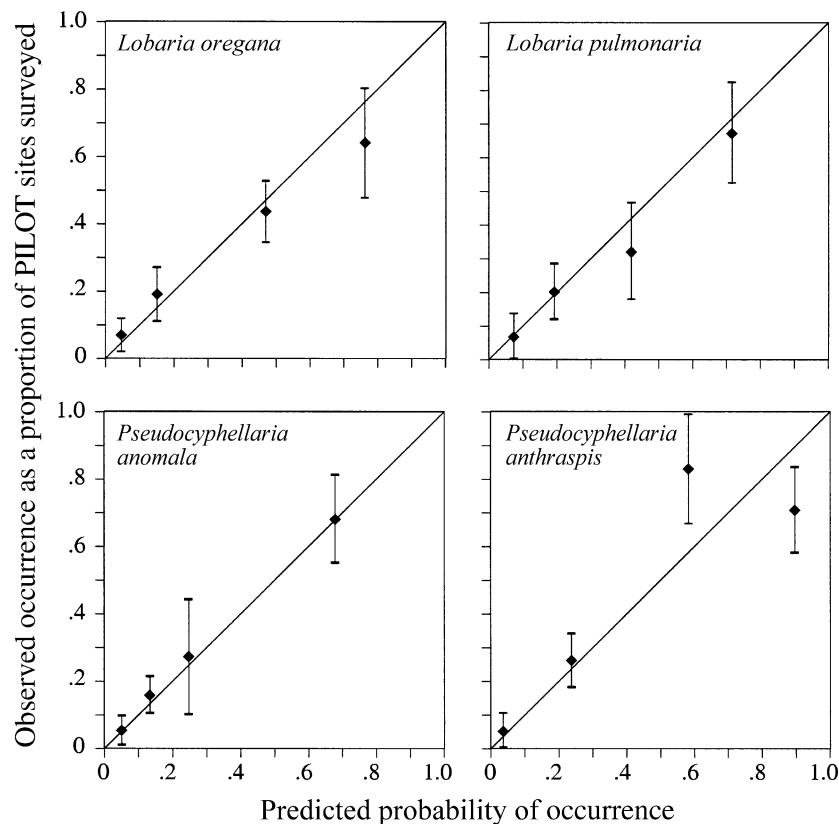


FIG. 3. Plot of the cumulative observed proportion of occurrences (plus 95% CI) against the probability of occurrence for four species of common lichen in the area of the Northwest Forest Plan. Predicted observations and 95% confidence intervals that encompass the expected 1:1 fit indicate models with high predictive capability.

evance to the distribution of the species being modeled. This shortcoming—the lack of strong linkage between species presence and ecologically relevant variables—is a common constraint faced by ecologists and cannot be satisfactorily resolved until spatially explicit predictions of such variables are more readily available (see Frescino et al. 2001).

There are several reasons why a species is or remains rare in a certain area, some of which can be captured by predictive models, while others cannot (see Gaston 1994 for an overview). Key patterns associated with rarity are (1) the main distribution of the species lies outside of the study area, ostensibly because the latter does not provide much suitable habitat (locally rare, otherwise abundant); (2) the species has a patchy, scattered distribution throughout its range, and populations are smaller and likely limited by competition and/or population constraints (generally and locally rare); (3) the species has a patchy distribution, but population size can be large at known sites (locally abundant, generally rare). Following the classification by Hanski (1982) and Collins et al. (1993), the first group would be classified as an “urban” type, the second as “rural” type, while the last group can be labeled as “satellite” type.

Accurate, predictive models for urban species that are dominant within their realized range are easier to construct than models for rural or satellite species. So-called urban species may not occur in many locations within the study area, but if the environment is suitable, they usually are present. Satellite type species may be restricted to few locations because there are actually only few suitable sites available. Often, however, dispersal limitations and other life history constraints limit this type from persisting, regenerating, or migrating to other suitable habitats. Predictive models may not easily predict the presences of satellite type species successfully, but the absences can be more accurately portrayed. This is because such species usually show a well defined ecological niche even though they do not occur on every suitable location. The most difficult group of species to model would be the rural type.

Because our approach uses associative models predicting rare from more common species, care must be exercised to ensure that the potential types of error are understood. For example, it is likely that rare species would have different niches than common species, by definition alone (Gaston 1994). In this case, the potential for commission error exists if the niche of the rare species is a subset of the more common species used

TABLE 4. Number of detections in the PILOT study area using simple random sampling (numerator), number of detections using the model-based stratifications (denominator), and model efficiency (value in parentheses).

Rare species	Common species			
	<i>Lobaria oregana</i>	<i>Lobaria pulmonaria</i>	<i>Pseudocyphellaria anomala</i>	<i>Pseudocyphellaria anthraspis</i>
<i>Lobaria scrobiculata</i>		13/26 (2.0)	13/36 (2.8)	
<i>Nephroma laevigatum</i>		19/23 (1.2)	19/48 (2.5)	19/60 (3.2)
<i>Nephroma occultum</i>		1/5 (5.0)	1/5 (5.0)	
<i>Nephroma parile</i>		7/14 (2.0)	7/16 (2.3)	
<i>Pseudocyphellaria rainierensis</i>	2/1 (0.5)	2/5 (2.5)	2/5 (2.5)	

Note: Model efficiency is estimated as the ratio of detections from the model-based stratification to those detected in the PILOT study area, Pacific Northwest Forest Plan. Empty cells indicate species that were never co-located.

to develop the stratification model. Although it is less likely that rare species' niches are larger than common species, it must be acknowledged that the potential for omission error also exists. Yet even though omission and commission errors occur with any predictive model, our use of models to develop sampling strata should overcome this problem. Any sampling based on models like ours should also include sampling effort allocated to the lowest probabilities of presence, thereby ensuring that sampling occurs across the spectrum of potential locations for the rare species. We acknowledge, however, that absences in the lower probability sites may still be confounded with low detection rates, such as is often the case for fungi (Dreisbach et al. 2002). We know of no simple resolution to this concern.

We argue that distinguishing different types of rarity may help to improve modeling of such rare events. To our knowledge, none of the cited studies for modeling rare species used any sort of formal model-based stratification design based on more common species. Wiser et al. (1998) demonstrated the difficulty of generating viable models from very limited observations only (10–24 presences), even though detailed, site-specific bio-geochemical predictors were additionally used to increase predictive capabilities. Engler et al. (2004) used two initial data sets of 46 and 77 occurrences of the rare species *Eryngium alpinum* L. to generate a model-based design improved their ability to identify new sites. A similar approach was performed by Spelduto and Congalton (1996). Such an approach is only a viable alternative if a minimal set of observations is available up-front for a realistic initial model. This was not the case for any of the tested rare species in our study.

A possible reason our effort worked so well is because all of our tested lichen species are cyanolichens, and they therefore have distinct and similar habitat requirements compared to green algal lichens. Green algal lichens, which are approximately 75% of the lichens in the Pacific Northwest, occupy wider ranges of habitats. Consequently the use of intrageneric species might not work so well in when applied to these lichens. In a study on coastal lichens, L. Geiser (*unpublished data*) documented the presence of *Ramalina farianacea* on nearly every plot, leading to an almost virtual over-

lap with two species, *R. thrausta* and *R. menziessi*, known to be rare. In such cases as this, it might be better to select a less common species that also has high association the targeted rare species, as we did here.

Our results suggest a model-based stratification design that predicts rare species from more common ones can improve detection likelihoods, especially if only very limited information is available. We expect that so-called "satellite" and "rural" type species would benefit from most from such a design, since their realized niches are difficult to detect if only very few observations are available for direct modeling an evaluation of the rare species. If one is interested in one to few rare species for management and evaluation, the best approach seems to be a combination of several approaches. Model-based stratification as presented here would logically be a first step in a sampling and modeling process.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of the following people in field collection, identification and database entry: S. Berryman, M. Boyll, C. Derr, A. Ingersoll, D. Glavich, K. Gossen, A. Mikulin, J. Riley, and R. Ulrich. Special thanks to P. Halonen and B. Ryan (deceased) for help with difficult identifications, and to the many others who assisted with field collections. Field work was funded by the PNW Region Air Program. We are also grateful for the assistance from a large number of individuals working on Survey and Manage species in the Pacific Northwest. Specifically, we thank T. Brumley, N. Diaz, B. Rittenhouse, and N. Middlebrook for guidance on Forest Service and Bureau of Land Management Survey and Manage Issues. Two anonymous reviewers and T. B. Murphy helped improve the manuscript with their insights. Funding was provided by the State of Oregon Bureau of Land Management through a cooperative research arrangement with the USGS Forest and Rangeland Ecosystem Science Center (FRESC), Corvallis, Oregon, and the USGS Utah Cooperative Research Unit, Utah State University. Last, we thank A. Ellison and A. Agrawal for organizing the special features section on rarity, and for inviting our participation.

LITERATURE CITED

- Alexander, H. M., N. A. Slade, and W. D. Kettle. 1997. Application of mark-recapture models to estimation of the population size of plants. *Ecology* 78:1230–1237.
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. Classification and regression trees. Wadsworth and Brooks/Cole, Monterey, California, USA.

- Collins, S. L., S. M. Glenn, and D. W. Roberts. 1993. The hierarchical continuum concept. *Journal of Vegetation Science* **4**:149–156.
- Cutler, R., T. C. Edwards, Jr., J. Alegria, and D. McKenzie. 2002. A sample design framework for Survey and Manage species under the Northwest Forest Plan. Proceedings of the Section on Statistics and Environment, 2001 Joint Statistical Meeting. American Statistical Association, Alexandria, Virginia, USA.
- De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* **81**:3178–3192.
- Dreisbach, T. A., J. E. Smith, and R. Molina. 2002. Challenges of modelling fungal habitat: when and where do you find chanterelles? Pages 475–481 in J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. Predicting species occurrence: issues of accuracy and scale. Island Press, Covello, California, USA.
- Edwards, T. C., Jr., R. Cutler, L. Geiser, J. Alegria, and D. McKenzie. 2004. Assessing rarity and seral stage association of species with low detectability: lichens in western Oregon and Washington forests. *Ecological Applications* **14**:414–424.
- Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* **41**:263–274.
- Esseen, P.-A., K.-E. Renhorn, and R. B. Pettersson. 1996. Epiphytic lichen biomass in managed and old-growth boreal forests: effect of branch quality. *Ecological Applications* **6**:228–238.
- Ferrier, S., and G. Watson. 1997. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. Environment Australia, New South Wales National Parks and Wildlife Service, Armidale, New South Wales, Australia.
- Frescino, T. S., T. C. Edwards, Jr., and G. G. Moisen. 2001. Modelling spatially explicit forest structural attributes using generalized additive models. *Journal of Vegetation Science* **12**:15–26.
- Gaston, K. J. 1994. *Rarity*. First edition. Chapman and Hall, London, UK.
- Geiser, L. 2004. Manual for monitoring air quality using lichens on national forests of the Pacific Northwest. Technical Paper R6-NR-AQ-TP-1-04. USDA Forest Service, Pacific Northwest Region, Portland, Oregon, USA.
- Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler. 2002. The national elevation data set. *Photogrammetric Engineering and Remote Sensing* **68**:5–12.
- Goerck, J. M. 1997. Patterns of rarity in the birds of the Atlantic forest of Brazil. *Conservation Biology* **11**:112–118.
- Green, R. H., and R. C. Young. 1993. Sampling to detect rare species. *Ecological Applications* **3**:351–356.
- Gregoire, T. G., H. T. Valentine, and G. M. Furnival. 1995. Sampling methods to estimate foliage and other characteristics of individual trees. *Ecology* **76**:1181–1194.
- Hanski, I. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* **38**:210–221.
- Hill, N. M., and P. A. Keddy. 1992. Prediction of rarities from habitat variables: coastal plain plants on Nova Scotian lakeshores. *Ecology* **73**:1852–1859.
- Hoeting, J. A., M. Leecaster, D. Bowden. 2000. An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological and Environmental Statistics* **5**:102–114.
- Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**:299–314.
- James, F. C., C. E. McCulloch, and D. A. Wiedenfeld. 1996. New approaches to the analysis of population trends in land birds. *Ecology* **77**:13–21.
- Jensen, M. E., and H. R. Haise. 1963. Estimating evapotranspiration from solar radiation. *Journal of Irrigation and Drainage Engineering-ASCE* **89**:15–41.
- Lohr, S. L. 1999. *Sampling: design and analysis*. Duxbury Press, Pacific Grove, California, USA.
- Maina, G. G., and H. F. Howe. 2000. Inherent rarity in community restoration. *Conservation Biology* **14**:1335–1340.
- Manly, B. F. J. 1997. *Randomization, bootstrap and Monte Carlo methods in biology*. Chapman and Hall, New York, New York, USA.
- Max, T. A., H. T. Schreuder, J. W. Hazard, J. Teply, and J. Alegria. 1996. The Region 6 vegetation inventory and monitoring System. General Technical Report PNW-RP-493. USDA Forest Service, Pacific Northwest Research Station, Portland, Oregon, USA.
- McCune, B. 1993. Gradients in epiphyte biomass in three *Pseudotsuga-Tsuga* forest of different ages in western Oregon and Washington. *Bryologist* **96**:405–411.
- Molina, R., D. Pilz, J. Smith, S. Dunham, T. Dreisbach, T. O'Dell, and M. Castellano. 2001. Conservation and management of forest fungi in the Pacific Northwestern United States: an integrated ecosystem approach. Pages 19–63 in D. Moore, M. M. Nauta, S. E. Evans, and M. Rotheroe, editors. Fungal conservation. Issues and solutions. Cambridge University Press, New York, New York, USA.
- Murphy, D. D., and B. R. Noon. 1992. Integrating scientific methods with habitat conservation planning: reserve design for northern spotted owls. *Ecological Applications* **2**:3–17.
- Olsen, A. R., and H. T. Schreuder. 1997. Perspectives on large-scale natural resource surveys when cause-effect is a potential issue. *Environmental and Ecological Statistics* **4**:167–180.
- Peterson, E. B., and B. McCune. 2001. Diversity and succession of epiphytic macrolichen communities in low-elevation managed conifer forests in western Oregon. *Journal of Vegetation Science* **12**:511–524.
- Philippi, T. 2005. Adaptive cluster sampling for estimation of abundances within local populations of low-abundance plants. *Ecology* **86**:1091–1100.
- Rosso, A. L., B. McCune, and T. R. Rambo. 2000. Ecology and conservation of a rare, old-growth-associated canopy lichen in a silvicultural landscape. *Bryologist* **103**:117–127.
- Sauer, J. R., and S. Droege, editors. 1990. Survey designs and statistical methods for the estimation of avian population trends. Biological Report 90(1), U.S. Fish and Wildlife Service, Washington, D.C., USA.
- Schreuder, H. T., T. G. Gregoire, and G. B. Wood. 1993. *Sampling methods for multiresource forest inventory*. John Wiley and Sons, New York, New York, USA.
- Sheldon, A. L. 1988. Conservation of stream fishes: patterns of diversity, rarity, and risk. *Conservation Biology* **2**:149–156.
- Shiver, B. D., and B. E. Borders. 1996. *Sampling techniques for forest resource inventory*. John Wiley and Sons, New York, New York, USA.
- Sillett, S. C., and M. N. Goslin. 1999. Distribution of epiphytic macrolichens in relation to remnant trees in a multiple-age Douglas-fir forest. *Canadian Journal of Forest Research* **29**:1204–1215.
- Sperduto, M. B., and R. G. Congalton. 1996. Predicting rare orchid (small-whorled Pogonia) habitat using GIS. *Photogrammetric Engineering and Remote Sensing* **62**:1269–1279.
- Stow, C. A., S. R. Carpenter, K. E. Webster, and T. M. Frost. 1998. Long-term environmental monitoring: some perspectives from lakes. *Ecological Applications* **8**:269–276.

- Thomas, J. W., E. D. Forsman, J. B. Lint, E. C. Meslow, B. R. Noon, and J. Verner. 1990. A conservation strategy for the northern spotted owl. Interagency Scientific Committee to Address the Conservation of the Northern Spotted Owl. Report No. 1990-791-171/20026. United States Government Printing Office, Washington, D.C., USA.
- Thornton, P. E., and S. W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agricultural and Forest Meteorology* **93**:211–228.
- Thornton, P. E., S. W. Running, and M. A. White. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology* **190**: 214–251.
- U.S. Department of Agriculture, Forest Service, and U.S. Department of Interior, Bureau of Land Management. 1994. Record of decision for amendments to Forest Service and Bureau of Land Management planning documents within the range of the northern spotted owl. (http://www.or.blm.gov/roseburg/ROD_RMP/roseburg/appendices/appendixa.html).
- U. S. Department of Agriculture, Forest Service, and U.S. Department of Interior, Bureau of Land Management. 2001. Record of decision and standards and guidelines for amendments to the survey and manage, protection buffer, and other mitigation measures. Standards and guidelines. (<http://www.or.blm.gov/nwfpnepa/FSEIS-2000/ROD-SandG.pdf>).
- Wiser, S. K., R. K. Peet, and P. S. White. 1998. Prediction of rare-plant occurrence: a southern Appalachian example. *Ecological Applications* **8**:909–920.

APPENDIX A

A table showing the expected number of detections on model-assisted sampling for *Pseudocyphellaria rainierensis* from the classification tree model of *Lobaria oregana* is presented in ESA's Electronic Data Archive: *Ecological Archives* E086-057-A1.

APPENDIX B

A table showing the expected number of detections based on model-assisted sampling for *Pseudocyphellaria rainierensis* and *Nephroma laevigatum* from the classification tree model of *Lobaria pulmonaria* is presented in ESA's Electronic Data Archive: *Ecological Archives* E086-057-A2.

APPENDIX C

A table showing the expected number of detections based on model-assisted sampling for *Nephroma occultum* and *Nephroma parile* from the classification tree model of *Pseudocyphellaria anomala* is presented in ESA's Electronic Data Archive: *Ecological Archives* E086-057-A3.

APPENDIX D

A table showing the expected number of detections based on model-assisted sampling for *Pseudocyphellaria anthraspis* is presented in ESA's Electronic Data Archive: *Ecological Archives* E086-057-A4.